Matching of omnidirectional and perspective images using the hybrid fundamental matrix

Luis Puig, Josechu Guerrero, Peter Sturm

Matching of omnidirectional and perspective images using the hybrid fundamental matrix

Luis Puig[1], J.J. Guerrero[1] and Peter Sturm[2]
lpuig@unizar.es, jguerrer@unizar.es, Peter.Sturm@inrialpes.fr

[1]DIIS-I3A, Universidad de Zaragoza, Zaragoza, Spain
[2] INRIA Rhône-Alpes, Montbonnot, France

Abstract. In this paper we present an automatic hybrid matching system mixing images coming from central catadioptric systems and conventional cameras. We analyze three models of hybrid fundamental matrices. We perform experiments with synthetic and real data to test the behavior of these three approaches. The sensitivy to noise of lifted coordinates induces to select the simplest model to build an automatic matching system between these kind of images. Scale invariant features with a simple unwarping tool are considered to help initial putative matching. Then a robust estimation gives an estimation of the hybrid fundamental matrix and allows to detect wrong matches. Experimental results show the feasibility of this system.

1 Introduction

Hybrid image matching is a way for establishing a relation between two or more images coming from different camera types. The combination of omnidirectional with perspective images is important since a single omnidirectional image contains a more complete description of the object or place it represents than a perspective image, which is the most regular visual acquisition system. Some areas where the combination of these cameras has an important role are localization and recognition, since a database of omnidirectional images would be more representative with less data, and perspective images are the simplest query images [1]. In camera networks omnidirectional images not only provide a good reference but also minimize the possibility of fatal occlusions in a tracking process. The perspective images capture more detail information in higher resolutions [2].

In the literature we find a common strategy to deal with the matching of pairs of uncalibrated images. It consists in the use of epipolar geometry. This geometry has already been developed for both kind of images. In the perspective case it has been studied for a long time and a good explanation can be found in [3]. The epipolar geometry for central catadioptric cameras has been developed by Svoboda and Pajdla [4]. Some authors have developed the hybrid epipolar geometry dealing with different catadioptric systems. Sturm [5] proposes two models of hybrid fundamental matrices, a 4×3 fundamental matrix to relate

a para-catadioptric view and a perspective view and a 6×3 fundamental matrix to relate a perspective view and a general central catadioptric view. In [1] Menem et al. propose an algebraic constraint on corresponding image points in a perspective image and a circular panorama. They use a lifting from 3-vector to 6-vector to describe Plücker coordinates of projected rays. In [6] Claus et al. propose the lifting of image points to 6-vectors to build a general purpose model for radial distortion in wide angle and catadioptric lenses. Recently Barreto and Daniilidis [7] propose a general model that relates any type of central cameras including catadioptric systems with mirrors and lenses and conventional cameras with radial distortion. They apply the lifted coordinates in both images. This lifted coordinates correspond to a map from $\wp^2$ to $\wp^5$ through Veronese maps. They propose a 6×6 fundamental matrix to compute the geometrical constraint. Notice that all the approaches mentioned above work with central catadioptric systems which are modeled as well as the perpective cameras by the sphere model, originally proposed by Geyer and Daniilidis [8] and modified by Barreto and Araujo [9].

All the approaches mentioned above need pairs of putative corresponding points between the views. These correspondences are built from previously detected relevant features. Perhaps the most used extractor is the SIFT [10]. However, if SIFT features extracted in an omnidirectional image are matched to features extracted in a perspective image the results are not good, this is because SIFT is scale invariant but not camera invariant [11]. We show that with a simple polar transformation applied in the omnidirectional image, SIFT points can still be useful. Note that this transformation does not require camera calibration, since only a symmetry of revolution and a 360° field of view are assumed.

For image matching, some simplified or approximate models are required, since they must be integrated into a RANSAC estimation technique to cope with a large number of mismatches. We perform experiments with synthetic and real data in order to analyze the behavior of the hybrid fundamental matrices approaches. We selected the one with a better performance to construct our hybrid matching system. The goal is not to have a perfect model for calibration and 3D modeling, but to obtain a reasonable model for robust matching from uncalibrated hybrid image sets, which up to our knowledge has not yet been developed. Also this approach can be used as a part of a more complex system where the properties of the hybrid fundamental matrix such as the epipoles can provide information about camera location or motion.

2 Hybrid Image Matching using Epipolar Geometry

In this section we explain the epipolar geometry between omnidirectional and perspective images. In [4] Svoboda et al. explain that in the catadioptric image epipolar lines become epipolar conics. When we mix perspective and omnidirectional images, a point in the perspective image is mapped to its corresponding epipolar conic in the omnidirectional image while a point in the omnidirectional image is mapped to its corresponding epipolar line in the perspective image. In

general the relation between omnidirectional and perspective images with the fundamental matrix that we call *hybrid fundamental matrix* is established by

$$\hat{\mathbf{q}}_c^T \mathbf{F}_{cp} \mathbf{q}_p = 0 \tag{1}$$

subscripts p and c denote perspective and catadioptric respectively.

We define the coordinate vectors in the two images, $\mathbf{q}_p = (q_1, q_2, q_3)^T$ is a point in the perspective image in homogeneous coordinates. To describe $\hat{\mathbf{q}}_c$ we use two different vector representations, depending on the shape of the epipolar conic. One is the general representation for any shape of epipolar conic which is a 6-vector. The other one is a special case where the shape of the conic is a circle and the coordinate vector is a 4-vector. These two representations are called the "lifted coordinates" of a point in the omnidirectional image.

As explained before, epipolar lines become epipolar conics in omnidirectional images. Conics can be represented in homogeneous coordinates as

$$c_1 q_1^2 + c_2 q_2^2 + c_3 q_3^2 + c_4 q_1 q_2 + c_5 q_1 q_3 + c_6 q_2 q_3 = 0 \tag{2}$$

Eq. (2) can also be seen as the product $\hat{\mathbf{q}}^T \mathbf{c} = 0$,

$$\begin{pmatrix} q_1^2 & q_2^2 & q_3^2 & q_1 q_2 & q_1 q_3 & q_2 q_3 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \end{pmatrix} = 0 \tag{3}$$

where $\hat{\mathbf{q}}$ is called the "lifted coordinates" of $\mathbf{q}$.

If the conic is obtained through the 6×3 fundamental matrix (F63) from a corresponding point $\mathbf{q}_p$, then $\mathbf{c} \sim \mathbf{F}_{cp} \mathbf{q}_p$.

Para-catadioptric system. The para-catadioptric system is a catadioptric system which is composed by a parabolic mirror and an orthographic camera. In this case the shape of the conic is a circle and we can simplify the parameterization of the conic. With only 4 parameters we can define the epipolar conic

$$c_1(q_1^2 + q_2^2) + c_2 q_1 q_3 + c_3 q_2 q_3 + c_4 q_3^2 = 0 \tag{4}$$

then the lifted coordinates are $(q_1^2 + q_2^2, q_1 q_3, q_2 q_3, q_3^2)^T$

So the point in the omnidirectional image is represented with 4 lifted coordinates $\hat{\mathbf{q}}$ and the fundamental matrix is 4×3 (F43) in such a way that $\mathbf{c} \sim \mathbf{F}_{cp} \mathbf{q}_p$.

Barreto-Daniilidis model. As we mentioned in the introduction there exists a theoretical work relating any kind of central cameras. One of these mixtures is the case of central catadioptric cameras using a hyperbolic mirror and pin-hole cameras. According to [7] this is the only case where there exists a fundamental matrix involving a catadioptric system with a hyperbolic mirror. The way to compute the fundamental matrix consists in lifting the coordinates of the points

in both images to a Veronese map, which is a mapping from $\wp^2$ to $\wp^5$ defined by the following operator

$$\mathbf{\Gamma}(\mathbf{q}, \dot{\mathbf{q}}) = (q_1\dot{q_1}, \frac{q_1\dot{q_2} + q_2\dot{q_1}}{2}, q_2\dot{q_2}, \frac{q_1\dot{q_1} + q_3\dot{q_1}}{2}, \frac{q_2\dot{q_3} + q_3\dot{q_2}}{2}, q_3\dot{q_3})^T \qquad (5)$$

where $\mathbf{q}$ and $\dot{\mathbf{q}}$ are 3-vector representing homogeneous coordinates of a point in an image. If we apply this operator to the same point then we obtain the lifted coordinates as described above.

This operator preservers homogeneity and is suitable to deal with quadratic functions because it discriminates the entire set of second order monomials [7]. Note that this lifting operator does not give exactly the same as Eq. 3 but essentially it is the same. It gives the same vector up to a permutation of its coefficients. The idea of this mapping is to obtain a bilinear relation between the two views, which is achieved with a 6×6 homogeneous matrix F66. This matrix works in a similar way as F63 from perspective to omnidirectional images. The difference is the opposite direction, because it relates a point in the omnidirectional image to a conic in the perspective. This conic is composed by two lines. These lines are the forward looking epipolar line and the backward looking epipolar line. To compute these lines we need to obtain the epipole in the perspective image which is achieved imposing rank 2 in the epipolar conic and then computing the null vector. This rank 2 imposition does not guarantee the good estimation of the epipole. For a more detailed explanation [7] can be consulted.

All these hybrid fundamental matrices works in two directions, from perspective to catadioptric images and vice-versa. The first direction relates perspective image points to epipolar conics $\mathbf{c} \sim \mathbf{F}_{cp}\mathbf{q}_p$ and the second, omnidirectional image points to epipolar lines $\mathbf{l}^T \sim \hat{\mathbf{q}}_c^T\mathbf{F}_{cp}$.

From these three approaches the theoretically correct is the F66 model which comtemplates different mirror shapes such as the hyperbolic and parabolic while the F43 model is only valid for parabolic mirrors. The F63 model is in between these two models being more general than the F43 model but not correct as F66 model.

2.1 Computation of the Hybrid Fundamental Matrix

We use a DLT-like approach [3] to compute the hybrid fundamental matrix. It is explained as follows. Given n pairs of corresponding points $\hat{\mathbf{q}}_c \leftrightarrow \mathbf{q}_p$, solve the equations $\hat{\mathbf{q}}_c^T\mathbf{F}_{cp}\mathbf{q}_p = 0$ to find $\mathbf{F}_{cp}$. The solution is the least eigenvector, $\mathbf{f}$ of $\mathbf{A}^T\mathbf{A}$, where $\mathbf{A}^T$ is the equation matrix (we omit the subindices c and p for clarity)

$$\mathbf{A}^T = \begin{pmatrix} \hat{\mathbf{q}}_1 q_{11} & \cdots & \hat{\mathbf{q}}_1 q_{1m} \\ \vdots & \ddots & \vdots \\ \hat{\mathbf{q}}_n q_{n1} & \cdots & \hat{\mathbf{q}}_n q_{nm} \end{pmatrix}. \qquad (6)$$

The number of pairs of corresponding points (n) needed to compute the hybrid fundamental matrix depends on the number of elements of the matrix, n = No. of elements of $\mathbf{F}$-1 (scale factor). Each pair of corresponding points gives one equation. As we are interested in the matching and simultaneous automatic computation of the fundamental matrix, a robust estimation is considered and is summarized as follows:

1. **Initial Matching**. Scale invariant features are extracted from perspective and unwarped omindirectional images (this process is explained in section 4.1) and matched based on their intensity neighborhood.
2. **RANSAC robust estimation**. Repeat for r samples, where r is determined adaptively:
   (a) Select a random sample of k corresponding points, where k depends on what model we are using (if F43, $k = 11$, if F63, $k = 17$ or if F66 $k = 35$). Compute the hybrid fundamental matrix $\mathbf{F}_{cp}$ as mentioned above.
   (b) Compute the distance d for each putative correspondence, d is the geometric distance from a point to its corresponding epipolar conic. This distance is explained below.
   (c) Compute the number of inliers consistent with $\mathbf{F}_{cp}$ by the number of correspondences for which $d < t$ pixels.
   Choose the $\mathbf{F}_{cp}$ with the largest number of inliers.
3. **Non-linear re-estimation**. Re-estimate $\mathbf{F}_{cp}$ from all correspondences classified as inliers by minimizing the distance in both images to epipolar conics and epipolar lines, using an unconstrained nonlinear optimization.

The $\mathbf{F}_{cp}$ is used to eliminate outliers which are those point correspondences for which $d > t$. Using this scheme the F43 has a great advantage over the F63 and the F66, since it just needs less correspondences. In practice, there is an agreement between the computational cost of the search in the space of solutions, and the probability of failure $(1-p)$. A random selection of r samples of k matches ends up with a good solution if all the matches are correct in at least one of the subsets. Assuming a ratio ε of outliers, the number of samples to explore is $r = \frac{\log(1-p)}{\log(1-(1-\varepsilon)^k)}$. For example using a probability $p = 99\%$ of not failing in the random search and 30% of outliers (ε), 231 iterations are needed to get a result. On the other hand, if we use the F63, 1978 iterations are needed for the same level of confidence. In the case of the F66 the number of iterations increases and becomes prohibitive for matching (1.2×10^6 are required).

2.2 Point to conic distance

The most suitable method to compute $\mathbf{F}$ is to minimize the distance to the conic in the omnidirectional image. We use the approach proposed by Sturm [12] where the point-to-conic distance computation is replaced by a point-to-point distance computation. He defines a cost function based on the geometric distance between a point in homogeneous coordinates and a conic $dist(\mathbf{q}, \mathbf{C})$, i.e., the distance between $\mathbf{q}$ and the point on $\mathbf{C}$ that is closest to $\mathbf{q}$. He uses a parameterization which includes a projective transformation $\mathbf{P}$. The parameterization guarantees

that we have points lying on a conic and that we can transform *proper* conics using a projective transformation $\mathbf{C}' \sim \mathbf{P}^{-T}\mathbf{C}\mathbf{P}^{-1}$, where $\mathbf{C}'$ is a *proper* conic. The parameterization is explained as follows. For a measured point $\mathbf{q}$, we parameterize a point $\bar{\mathbf{q}}$, such that $\bar{\mathbf{q}}$ lies on a conic. The simplest way to do so is to choose the unit circle as support, in which case we may parameterize the $\bar{\mathbf{q}}$ by an angle α then $\bar{\mathbf{q}} = (\cos\alpha, \sin\alpha, 1)^T$. Finally the problem to be minimized is:

$$\min_{\alpha} dist(\mathbf{q}, \mathbf{P}\bar{\mathbf{q}})^2 \tag{7}$$

Once the function to be minimized is defined a unidimensional non-linear optimization process is performed to compute the point to conic distance in the omnidirectional image.

2.3 Rank 2 property

If the task we are interested in requires the epipoles of the fundamental matrix, it is mandatory to have a rank 2 matrix. Such applications are related to localization and motion. An example of these applications is the computing of the position of the perspective camera observed in the omnidirectional image which is given by the right epipole.

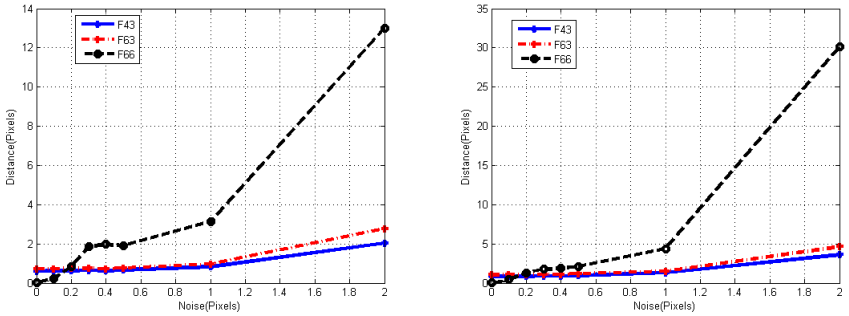
To deal with this problem we tried two options. One is to enforce this constraint minimizing the Frobenius norm using SVD as explained in [3] which we call *direct imposition*(DI). The other option is to perform a non-linear re-estimation process minimizing either the distance from points in one image to their corresponding epipolar conic or line in the other using the Levenberg-Marquardt (LM) algorithm. To guarantee the rank 2 we use a matrix parameterization proposed in [13] which is called the *orthonormal representation* of the fundamental matrix. We consider the SVD of the estimated fundamental matrix $\mathbf{F} \sim \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. Since $\mathbf{F}$ must be rank 2, $\mathbf{\Sigma}$ should be equal to $diag(\sigma_1, \sigma_2, 0)$, where $\sigma_1 \geq \sigma_2 > 0$. We can scale $\mathbf{\Sigma}$ such that $\mathbf{F} \sim \mathbf{U}diag(1, \sigma, 0)\mathbf{V}^T$, where $\sigma = \sigma_2/\sigma_1$ ($\sigma_1 \neq 0$ since $\mathbf{F} \neq 0$) and $1 \geq \sigma > 0$. The fundamental matrix is recovered as

$$\mathbf{F} \sim \mathbf{u}_1\mathbf{v}_1^T + \sigma\mathbf{u}_2\mathbf{v}_2^T, \tag{8}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the columns of $\mathbf{U}$ and $\mathbf{V}$ respectively. This approach is originally applied to $O(3)$ matrices and easily adapted to F43 and F63. We also extend this approach to F66 matrix.

3 Hybrid Epipolar Geometry Experiments

In this section we present some experiments performed with synthetic data in order to analyze the behavior of the three fundamental matrices. We use a simulator which generates omnidirectional images coming from a catadioptric system using a hyperbolic mirror and perspective images from a pin-hole model. The two sensors are placed in a virtual volume of $5 \times 2.5 \times 7$ m. width, height and depth, respectively, where points are located randomly ($m \gg 35$). The perspective camera has a resolution of 1000×1000 pixels and is located at the origin of the coordinate system. The omnidirectional camera is located at $(0.5, 0.5, 3.5)$



(a) (b)

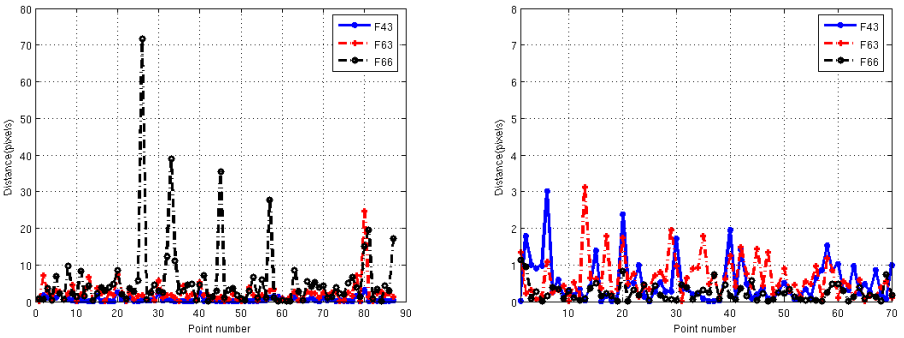
Fig. 1. Behavior of the three fundamental matrices in function of image noise(σ): Mean distances from points to epipolar conics in (a) omnidirectional image, and (b) perspective image.

which is close to the center of the scene. In this position we can have a good view of the whole scene. We use the sphere model [7] to generate the omnidirectional image. We take as reference a real hyperbolic mirror with semiaxis $a = 28mm$ and $b = 23mm$ and a diameter $60mm$. With these parameters and the sphere model we compute the mirror parameter $\xi = 0.9662$. As a common practice and because we are using lifted coordinates we apply a normalization to the image coordinates where the origin is the image center and the width and height are 1.

Once the points are projected. We add Gaussian noise characterized by its standard deviation to the coordinates of the projected points in both images. The fundamental matrices are computed using **fminsearch** function (provided by Matlab), minimizing the geometric distance from points in the images to epipolar conics [12]. For every σ we repeat the experiment 10 times in order to avoid particular cases due to random noise and the mean of these iterations is shown. Fig. 1 shows the distances from points to their corresponding epipolar conics and lines in function of image noise.

It is surprising that the non-exact theoretical model F43 shows a better performance in terms of residuals than F63 and F66 models. From Fig. 1 we can observe that when there is no noise present in the image the F66 shows the best performance in both directions, which is expected since F66 is the theoretically correct model. This changes when noise increases. In this case the F43 and F63 show a better performance, being consistent with the noise present in the images. The residuals of F63 are slightly larger than the ones from F43. We can say that F66 is instable when noise is present in the images. This behavior can be caused by the over-parameterization of the matrices, the more the parameters the higher the sensitivity to noise; that can also explain the difference between the F63 and F43.

As we mentioned before the difference between the F43 and F63 in a RANSAC approach is very important. Now we can see that F43 has some advantages over the F63 and F66 to solve the matching problem.



(a) (b)

Fig. 2. RMSE error from points to their corresponding epipolar conic using noiseless points. (a) Using the direct imposition of rank 2. (b) Using the LM algorithm.

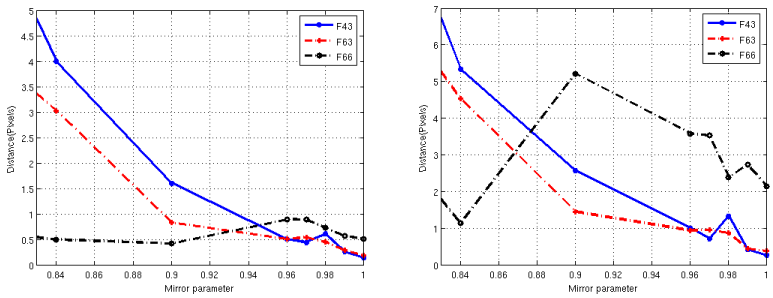
	True	F66		F63		F43	
	Value	DI	LM	DI	LM	DI	LM
x	467.38	467.89	469.31	461.22	460.55	467.8	467.39
y	728.37	728.41	727.83	716.55	724.42	728.41	727.67
D2C	0.0	7.2726	0.2932	2.4464	0.4920	0.4899	0.5015

Table 1. Epipoles estimated by the three fundamental matrices in the omnidirectional image. DI = direct imposition. LM = Levenberg-Marquardt. D2C = distance to epipolar conic.

First we present the estimation of the epipoles from the 3 approaches we have been analyzing. In this experiment we use noiseless data just to observe the behavior of these approaches to get a rank 2 matrix and then to compute the epipoles. We evaluate the performance of these approaches by the accuracy of the estimated epipoles and by the residual, which is the RMSE of the distances from the points used to compute the fundamental matrix to their corresponding epipolar lines and conics. In Fig. 2 we show the residuals for the three approaches imposing the rank 2 constraint by the *direct imposition* and by using the LM algorithm with *orthonormal representation*.

We can observe that some inaccuracies can appear if we apply the direct imposition, as in the case of point 26 with an error close to 70 pixels. This is explained because we are transforming a good solution given by the linear method into a new matrix with the rank 2 property but which doesn't minimize the distances between points and epipolar lines and conics. If we use the LM algorithm with the *orthonormal representation* we are imposing the rank 2 property and minimizing the distance between points and epipolar lines and conics using this new matrix.

Table 1 shows the epipoles from these two approaches. We can see from it that the three approaches give similar results in computing the epipole but we also observe an increment in the distance from points to conics and the minimization obtained with LM algorithm, all this as expected. Once more F43 shows an



(a) (b)

Fig. 3. Behavior of the three fundamental matrices in function of mirror parameter(ξ): Median distances from points to epipolar conics in (a) omnidirectional image, and (b) perspective image.

interesting behavior giving a small distance to conic even with the DI approach. This adds one advantage to F43.

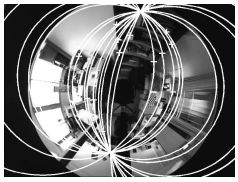
As observed from last experiments F43 shows a good performance dealing with images coming from a hyper-catadioptric system. In order to test this behavior we designed the following experiment. We modify the mirror parameter ξ from the hyperbolic case ($0 < \xi < 1$) to the parabolic case ($\xi = 1$) [7]. We add $\sigma = 0.5$ pixels Gaussian noise in both images and repeat the experiment 10 times to avoid bias since we are using random noise. In Fig. 3 we observe that F43 can deal with hyper-catadioptric images when the mirror shape is close to a parabola ($\xi = 1$) but when the mirror shape is more hyperbolic, in our experiment $\xi = 0.8$, F43 cannot deal with these omnidirectional images. On the other hand, the two more general models F66 and F63 which are designed to deal with these mirror shapes give better results.

3.1 Experiments with real images

We also performed experiments with real images coming from a catadioptric system using a hyperbolic mirror and from a conventional camera. The purpose of this experiment is to show the performance of these approaches to compute the hybrid epipolar geometry in real images. We use 70 manually selected corresponding points to compute them. In order to measure the performance of $\mathbf{F}$ we calculate the geometric error from each correspondence to their corresponding epipolar conic and line. We compute the root mean square of these two distances. Table 2 shows these distances for the no rank 2 matrix and for the two ways to obtain the rank 2 fundamental matrix. We can observe from this table that with rank 2 matrices error increases considerably. When we impose the rank 2 constraint we avoid a few degrees of freedom of the model (non-rank 2) that better adjusts to the data so residual error must be worse actually. From Fig. 4(c) we can observe that error in perspective image is bigger than in the omnidirectional one because some epipolar lines have been bad estimated from the epipolar degenerate conic. We have also observed that a great number of correspondences, bigger than the minimum is required. Using F63 we obtain good results having

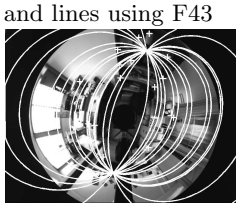
	Distance to epipolar conic			Distance to epipolar line		
	F43	F63	F66	F43	F63	F66
No Rank 2	0.6487	1.1942	1.0923	1.0802	1.3569	12.5677
Direct Imposition	3.1164	20.1165	96.9905	6.1821	32.2327	17.9322
Levenberg-Marquardt	0.7413	1.5734	14.4980	1.2684	2.7171	16.5136

Table 2. Mean of the distances to epipolar conics and lines for the 70 corresponding points in real images.



(a) Epipolar conics and lines using F43

(b) Epipolar conics and lines using F63



(c) Epipolar conics and lines using F66

Fig. 4. Experiments with real images using the three approaches of hybrid fundamental matrices.

50 (three times the minimum) correspondences. This gives a good reason to use the F43 to compute the hybrid matching.

4 Automatic Matching

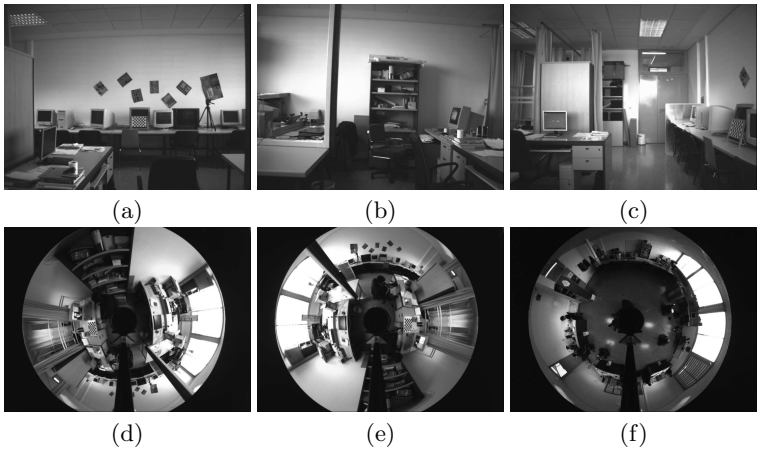
The first step of the matching process is to obtain an initial or putative set of pairs of corresponding features. It has been reported [14] reasonable matching of two omnidirectional images using well-known features like SIFT. But, as mentioned before SIFT is scale invariant but not camera invariant, making impossible to directly match omnidirectional images with perspective images using standard SIFT features.

4.1 Unwarping tool

We have observed that if we unwarp the omnidirectional image and get SIFT points from this image it is possible to obtain good matches between the omnidirectional and perspective image.

The unwarping tool is performed as follows:

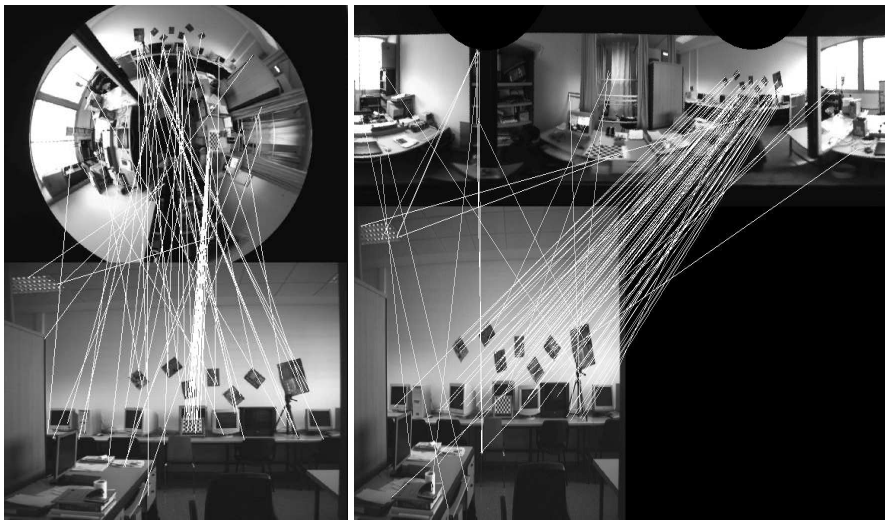
1. Using the center of the omnidirectional image a transformation from Cartesian to polar coordinates is computed, where $\rho = \sqrt{x^2 + y^2}$ (the radial component) represents the columns in the unwarped image and $\theta = \arctan\left(\frac{y}{x}\right)$ (the angular component), the rows.



(a) (b) (c)

(d) (e) (f)

Fig. 5. Some of the images used to test the automatic approach.



(a) (b)

Fig. 6. (a) Matching directly the SIFT points in the omnidirectional and perspective images. (b) Matching the unwarped omnidirectional image with the perspective image.

2. As the unwarped image has a higher resolution than the omnidirectional one, an interpolation process is needed in order to fill up the empty pixels. We choose bilinear interpolation to overcome this problem.

Note that we don't generate perspective views by this procedure. We generate views which are closer to perspective than the original omnidirectional images. This transformation does not require camera calibration. After SIFT points are extracted in the unwarped images, we then consider their coordinates in the

	SIFT points	Matches/Inliers
Omni	1991	68/0
Unwarped	2254	104/75

Table 3. Output from the SIFT matching using the omnidirectional and the unwarped omnidirectional image.

	Omni SIFT	Persp SIFT	Initial Matches (inliers/outliers)	Robust Epipolar Geometry matches(inliers/outliers)
Experiment 1	2219	856	28/21	23/0
Experiment 2	2247	1064	70/20	61/2
Experiment 3	2554	914	37/27	32/3
Experiment 4	1709	649	38/29	28/7
Experiment 5	1709	785	37/20	29/6

Table 4. Numerical results of the matches using the set of images.

original images, and compute epipolar geometry directly between the original input images.

4.2 Results

In this section we present experiments performing the automatic matching between omnidirectional and perspective images. First, we want to show the direct matching between SIFT points from omnidirectional (no unwarped) and perspective images. This can be seen in Fig. 6(a). The inliers and outliers obtained were counted manually. Table 3 shows that all the matches are wrong. Using the **unwarping tool** we repeat the experiment. The number of extracted features in the perspective image is 941. The resulting matching between these images is shown in Fig. 6b. Table 3 shows that now an important number of correct matches has been obtained.

Note that this initial matching between the perspective and the unwarped omnidirectional image has a considerable amount of inliers but also many outliers. This scenario requires a robust estimation technique and a geometric model like the hybrid epipolar geometry to detect the inliers and reject the outliers.

Four omnidirectional images and six perspective images are used to perform the following experiment. We use the algorithm explained in section 2.1 with the simplest model of hybrid fundamental matrix F43. We avoid the rank 2 constraint since we are just concerned about the matching problem and the epipoles are not needed. As explained before one of the advantages of F43 is that it requires fewer correspondences and as consequence fewer iterations in the RANSAC approach. Table 4 summarizes the results of this experiments giving the quantity of inliers and outliers in the initial and the robust matching. For example, in Experiment 1 we use images Fig. 5(b) and Fig. 5(d). The initial matching gives 57% of inliers. After applying the robust estimation we obtain 100% of inliers. Notice that just 5 inliers have been eliminated. Fig. 7 and Fig. 8 show two matches between omnidirectional and perspective images. The results show that the epipolar geometry eliminates most of the outliers.

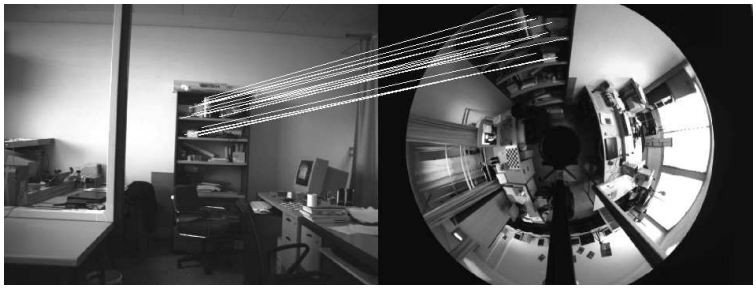


Fig. 7. Matching between omnidirectional and perspective image using the unwarping tool and the hybrid epipolar geometry.

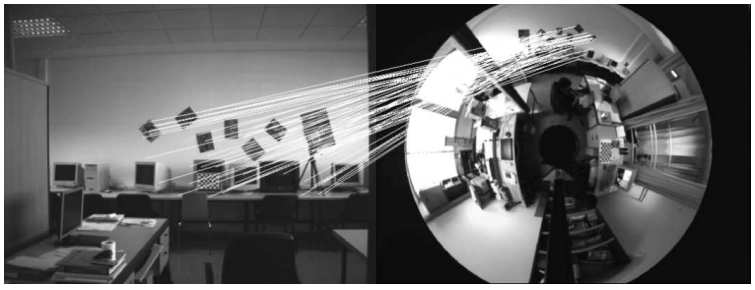


Fig. 8. Matching between omnidirectional and perspective image using the unwarping tool and the hybrid epipolar geometry.

5 Conclusions

In this work we have presented an automatic wide-baseline hybrid matching system using uncalibrated cameras. We performed experiments using synthetic and real data with three different approaches to compute the hybrid fundamental matrix, the perfect theoretical model F66, a general model F63 and a simplified model F43 were considered. We showed that the F66 model is very sensitive to the presence of noise in the images. Less sensitive are the F63 and F43 in decreasing order. We can say that this sensitivity is related in part to the overparameterization of the hybrid fundamental matrix. There exist some other factors that can cause the bad behavior of this theoretically correct model, such as the bad distribution of the matches in both images and the number of matches as exposed in experiments section. On the other side the F43 model has demonstrated to be useful in order to require fewer correspondences and as consequence fewer iterations to compute a robust fundamental matrix, having equal or even better performance than the other two approaches in presence of noise. We test these approaches in a simulator and the results obtained were congruent with the ones obtained from real images. We also prove that an easy polar transformation can be a useful tool to perform a basic matching between omnidirectional and perspective images. Finally the robust automatic matching proved its efficiency to match an omnidirectional image and a perspective image, both uncalibrated.

However there is still work to do, for example, it is needed a way to ponderate the error measure in the omnidirectional and perspective images, even these two measures represent a geometric distance they are not equivalent because of the non-homogeneous resolution of the omnidirectional image. Another important thing to develop is an adapted feature descriptor being camera invariant.

Acknowledgements

The authors thank the support given by the projects DPI2006-07928, UZ2007-TEC05 and DGA(CONSI+D)/CAI.

References

1. Menem, M., Pajdla, T.: Constraints on perspective images and circular panoramas. In Andreas, H., Barman, S., Ellis, T., eds.: BMVC 2004: Proceedings of the 15th British Machine Vision Conference, London, UK, BMVA, British Machine Vision Association (2004)
2. Chen, D., Yang, J.: Image registration with uncalibrated cameras in hybrid vision systems. In: WACV/MOTION. (2005) 427–432
3. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN: 0521623049 (2000)
4. Svoboda, T., Pajdla, T.: Epipolar geometry for central catadioptric cameras. Int. J. Comput. Vision **49** (2002) 23–37
5. Sturm, P.: Mixing catadioptric and perspective cameras. In: Workshop on Omnidirectional Vision, Copenhagen, Denmark. (2002) 37–44
6. Claus, D., Fitzgibbon, A.W.: A rational function lens distortion model for general cameras. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2005) 213–219
7. Barreto, J.P., Daniilidis, K.: Epipolar geometry of central projection systems using veronese maps. In: CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, IEEE Computer Society (2006) 1258–1265
8. Geyer, C., Daniilidis, K.: A unifying theory for central panoramic systems and practical applications. In: ECCV (2). (2000) 445–461
9. Barreto, J.a.P., Araujo, H.: Geometric properties of central catadioptric line images and their application in calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005) 1327–1333
10. Lowe, D.: Distinctive image features from scale-invariant keypoints. In: International Journal of Computer Vision. Volume 20. (2004) 91–110
11. Sturm, P., Ramalingam, S., Lodha, S.: On calibration, structure from motion and multi-view geometry for generic camera models. In Daniilidis, K., Klette, R., eds.: Imaging Beyond the Pinhole Camera. Volume 33 of Computational Imaging and Vision. Springer (2006)
12. Sturm, P., Gargallo, P.: Conic fitting using the geometric distance. In: Proceedings of the Asian Conference on Computer Vision, Tokyo, Japan, Springer (2007)
13. Bartoli, A., Sturm, P.: Non-linear estimation of the fundamental matrix with minimal parameters. IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004) 426–432
14. Murillo, A.C., Guerrero, J.J., Sagües, C.: Surf features for efficient robot localization with omnidirectional images. In: 2007 IEEE International Conference on Robotics and Automation, Roma. (2007)